

Ishan Madhuranga

+353 83 024 2827 | contact@excellenttorch.com | linkedin.com/in/excellent-torch | [Portfolio](#)

EDUCATION

Dublin Business School

Master of Science in Artificial Intelligence

Dublin, Ireland

Jan 2023 - Jan 2025

Solent University

Bachelor of Science in Cyber Security Management

Southampton, UK

April 2021 - April 2022

EXPERIENCE

AI Gameplay Programmer Intern

Sleepless Inc.

Sept 2024 – June 2025

Dublin, Ireland

- Developed and optimized AI behaviors and interactive elements, enhancing user experience overall.
- Debugged and optimized game performance for efficiency and stability, reducing system runtime by 60%.
- Stayed up to date with the latest Unreal Engine advancements and best practices, ensuring the latest feature integration.

Artificial Intelligence Engineer Intern

British College Of Applied Studies

Nov 2021 – Nov 2022

Colombo, Sri Lanka

- Contributed to academic research projects involving GANs, autonomous systems, and reinforcement learning, improving model performance overall.
- Built and deployed RESTful APIs using FastAPI and Flask, reducing system runtime.
- Documented all development processes, including model training, evaluation, and deployment strategies, in line with academic and industry standards.

Freelance AI/ML Specialist

Fiverr & Direct Contracts

Aug 2019 – Feb 2025

Colombo, Sri Lanka

- Delivered AI and ML solutions to clients via Upwork/Fiverr and direct contracts, leveraging Python to build and deploy models for deep learning, computer vision, and data analytics tasks.
- Conducted model optimization and integration for real-time applications and IoT/Client-based devices, enhancing overall system performance.
- Developed bespoke end-to-end technical solutions, including data preprocessing and model deployment using Amazon ECS/Azure Kubernetes, and tailored every minor adjustment for client satisfaction.

PROJECTS

Fast Neural Style Transfer In a 3D Environment in Real Time | *Python, C++*

June 2024 – Sept 2024

- Developed a custom transformer model for a real-time neural style transfer system for 3D environments
- Implemented intelligent NPC behavior using ConvAI for conversations with a custom fine-tuned BART LLM model with a RAG pipeline running on the backend to give unique personalities.

Optimized Inference Engine for ARM/Mobile Devices | *C++, ONNX, TensorRT*

Sept 2024

- Optimized deep learning models for real-time performance on ARM/Mobile devices by converting PyTorch models to ONNX and accelerating inference with TensorRT.
- Applied quantization and pruning techniques to shrink model size and cut latency, achieving faster and more efficient execution on NVIDIA Jetson hardware.
- Packaged the solution in Docker with automated tests and monitoring tools, ensuring reliability and scalability for production like deployments.

AI Research Assistant with RAG & LLM Integration | *Python, React, TypeScript*

Jan 2024 – April 2024

- Built an end-to-end AI solution with a RAG pipeline (LangChain + FAISS) and integrated LLMs for multi-turn reasoning, summarization, and document analysis.
- Developed and deployed a FastAPI backend with a React + TypeScript dashboard, enabling real-time, client-ready interaction. Optimized the pipeline for production environments.

TECHNICAL SKILLS

Languages: Python, C++, C#, SQL, JavaScript/TypeScript, HTML/CSS

Frameworks: React, Node.js, Transformer.js, LangChain, Flask, FastAPI, PyTorch, Tensorflow, MATLAB

Developer Tools: Git, Docker, Google Cloud Platform, AWS, VS Code, Visual Studio, Jupyter Notebooks, Oracle

Libraries: OpenCV, OnnxRuntime, Tkinter, pandas, NumPy, Threading, Matplotlib